CA Journal of Cultural Analytics

Measuring Modernist Novelty

Laura B. McGrath, Devin Higgins, and Arend Hintze

11.09.18

One may as well begin with two overly-cited quotations about novelty.[1] The first: "On or about December, 1910, human character changed." And, more predictably: "Make It New!"

That these two statements about change and innovation have been so frequently cited reveals a contradiction in modernist studies: novelty is commonplace, both for modernist writers and for literary critics. Novelty has become so firmly fixed as a central framing narrative for literary modernism that the two terms—"novelty", "modernism"—are often treated as synonymous. The period (roughly, 1890-1945) has been defined by its capacities for rupture and the rejection of tradition: a "paradigmatic shift, a major revolt;" " a ruthless break with any or all preceding historical conditions […] characterized by a never-ending process of internal ruptures and fragmentations within itself;" "a

---

[1] The authors would like to thank Thomas Padilla for his early work on this project. The preliminary study of the Bloom Filter's application to literary texts was conducted by Hintze, Higgins, and Thomas Padilla, and presented at DH2015 as "Patterns of Novelty in Literary Data."

powerful vortex of historical conditions that coalesce to produce sharp ruptures from the past [...] shattering change."[2] These revolts, breaks, and ruptures coalesced into the definitive sense of newness that would come to define the age for modernists and modernist literary critics alike, producing an exhaustive, exhausting commitment to newness in all its forms:

> The new spirit, the new form, the new reality, the new object, new facts, new organisms, new life, new values, a new world, new cosmos, new society, new culture or civilization, new era or epoch, new time and new age, to which the new consciousness and enhanced susceptibility would respond with a new sensitivity, a new corporeality, a new psychology aroused by new sensations; and these would give rise, naturally, to a new aesthetics, a new beauty —a new realism and a new archaism, enforced by a new rhythm—arising from new means and new methods: a new art, architecture, plastic expression, poetry, literature, new words and new language, music, drama, new optic realms, a new vision, a *new*. . . . The alacrity required to dispel the old order was audacious and impulsive, but it could always be called New.[3]

Then as now, the "new" means everything and nothing, as Jed Rasula's catalog aptly demonstrates. Novelty is so capacious a concept that it is a challenge to operationalize and measure—the very processes that would provide a deeper understanding of both the concept and its significance in modernism, literary history, and beyond. In this essay, we propose a method to measure one type of literary novelty—not distinct to, but emblematic of, high modernism: intratextuality. Situating modernism's literary innovations in the context of contemporaneous linguistic and scientific theory, Michael Levenson coined the term "intratextuality" to describe modernism's "aesthetic of composites" in his essay "Novelty, Modernity, Adjacency. [4] In keeping with Levenson's scientifically-inflected definition of modernist novelty, we derive our methods from computational biology and metagenomics, applying a Bloom Filter to the study of literary texts.

"Novelty" is a concept that has fascinated scientists, philosophers, and mathematicians as much as literary critics, as Michael North has shown; it is a field-specific term, taking on different meanings and tonal registers in different disciplines.[5] While literary critics such as Rasula may treat novelty as "new words and

---

[2]Astradur Eysteinsson, *The Concept of Modernism* (1990): David Harvey, "Modernity and Modernism," *Modernism,* ed. Tim Middleton (2003): 269; Susan Stanford Friedman, "Periodizing Modernism," *Modernism/modernity* 13.3 (September, 2006): 433.

[3]Jed Rasula, "Make it New." *Modernism/modernity* 17.4 (November, 2010): 713-733.

[4]"Michael Levenson,"Novelty, Modernity, Adjacency." *New Literary History* 42.4 (2011): 663-680.

[5]See North, Michael. *Novelty: A History of the New* (2015).

a new language," metagenomicists, for instance, understand novelty as a genetic mutation. A number of methods have been developed in bioinformatics in order to detect, measure, and analyze these instances of genetic novelty—methods that, when applied to literary texts, grant us a means of tracking novelty at its most microscopic level: the individual character. To that end, the aims of this essay are threefold: first, to demonstrate the potential of metagenomic methods for the analysis of intratextual novelty; second, to validate the measurement through close reading; third, to propose provisional hypotheses about the basic shapes of intratextual novelty that emerge from our preliminary results. Our hope is that this early analysis—a sort of methodological provocation—might open new possibilities for question, conversation, and collaboration.

* *

Methods: Intratextual Novelty and The Bloom Filter

Modernist novelty has been described as encompassing, but irreducible to "formal experiment, dislocation of conventional syntax, radical breaches of decorum, disturbance of chronology and spatial order, polysemy, obscurity, mythopoeic allusion, primitivism, irrationalism, structuring by symbol and motif rather than by narrative or argumentative logic"—to say nothing of the novelties of mass culture: advertisements, paperbacks, and pulps. A number of formal literary innovations were often revisions of past forms, such as Eliot's reliance on myth, rather than wholly new artistic experiments. A number of contradictions, then, rest at the heart of modernist novelty: both a rupture of epistemic newness and dime store tchotchkes, at once avant-garde and culture industry, equally innovation and renovation.

Beyond the definitional, novelty raises the problems of context and scale —that is, "new" and "not new" are constantly moving targets. "Novelty" is a comparative, contextual evaluation, judged only in relationship to the typical, the commonplace, the conventional, or the familiar. With each appearance of the new, both convention and the potential for innovation shift,[6] a problem that T.S. Eliot con-

---

[6]Novelty is not only always relative to what has come before it, but it also lays the groundwork for whatever novelty may surpass it, rearranging what scientists have called "the adjacent possible." Recent research attempting to test the intuition that one novelty begets another, has shown that the concept of novelty (describing a formation that is new only from a particular perspective, e.g. in the reader's experience moving page by page through a novel) is strongly linked to the concept of innovation (describing a formation that is new to all perspectives, e.g. to all readers in relation to their larger reading lives). Both forms of newness, local and global, appear to align with the same mathematical models whereby if "a novelty of any kind occurs, it does not occur alone. It comes

sidered at length in his seminal essay, "Tradition and the Individual Talent": "The existing monuments form an ideal order among themselves, which is modified by the introduction of the new (the really new) work of art among them… for order to persist after the supervention of novelty, the *whole* existing order must be, if ever so slightly, altered; and so the relations, proportions, values of each work of art toward the whole are readjusted."[7] For Eliot, novelty's comparative nature is also a problem of scale, in which the context of the "whole existing order" is both ever-shifting *and* ever-expanding. Eliot's "new (the really new)" would be methodologically impossible to measure.

In his essay, "Novelty, Modernity, Adjacency," Michael Levenson advances a heuristic for understanding novelty-in-context: intratextuality. Instead of the revolutionary novel, Levenson proposes intratextuality as model of the *relationally* novel, developed in response to both Bertram Russell's logical atomism and Saussurean linguistics. Levenson writes, "Intratextuality designates all those relations threaded within the boundaries of the artifact itself, all the mirroring and summoning of one piece of text by another… the tonal play that aligns distant elements without building narrative connections…The persistent impulse to say it again as said before, to repeat and resume, to gather up the foregoing words of the text, to have them melt into one another… assemble[d] into an echo chamber of mutual relations." [8] The result is an "aesthetic of composites, multivoiced, polyphonic compounds," representative of the usual high modernist suspects — "The Waste Land," *Ulysses,* montage, Cubism. Modernism's inheritance, Levenson argues, is the "inveterate relation of novelty and context," made particularly apparent by moments of intratextuality. Intratextual novelty is not an absolute novelty, in which individual work readjusts an Eliotic "whole existing order," but rather, novelty contained within a text—the means by which an individual text might evince novelty within itself. Delimiting novelty's context to the boundaries of a text provides a much more stable basis for sorting the "new" from the "not new," and an eminently more measurable scale. It is here that we begin, adopting Levenson's concept of intratextuality as a means of operationalizing novelty as a high modernist form.

To identify moments of novelty within the context of the familiar, we turn to bioinformatics, a field that poses similar questions about novelty vs. familiarity - not about literature, but about genetics, studying the progression of genetic se-

---

with an entourage of surrounding possibilities, a cloud of other potentially new ideas or experiences that are thematically adjacent to it and hence can be triggered by it." See F. Tria, V. Loreto, V. D. P. Servedio, and S. H. Strogatz. "The Dynamics of Correlated Novelties." *Nature* (July 2014).

[7] T.S. Eliot and Frank Kermode, *Selected Prose of T.S. Eliot.* (New York: Farrar, Straus, and Giroux, 1988), 38.

[8] Michael Levenson, "Novelty, Modernity, Adjacency." *New Literary History* 42.4 (2011): 668-70.

quences for previously unseen materials.[9] Novelty, in this context, means that a particular DNA sequence either deviates subtly from an earlier found sample or has never been seen before. In order to identify novelty within samples, metagenomicists employ a probabilistic, memory-efficient data structure called the Bloom Filter, which trades an acceptable amount of error for enough data capacity to work at scale.[10] Simply put, the Bloom Filter determines whether or not an element is a part of a set, allowing scientists to quickly identify moments of genetic variance for further study.[11] When applied to literature, the Bloom Filter makes it possible to computationally identify moments of heightened or diminished newness within a text ("has this string of text been seen before?"), as well as the degree of internal variation or novelty that occurs across the text as a whole ("how many times?"). Instead of an externally defined comparison ("the whole existing order"), the Bloom Filter analyzes character-sequence novelty within a text itself, providing us a measurement of intratextual novelty.

The Filter scans the entirety of a text, tracking the repetition of language at incredibly small, precise intervals called *k*-mers; in our case, a fixed 12-character window called a 12-mer. The Bloom Filter evaluates the text one k-mer at a time, starting at the beginning then advancing one character, such that 12-mer 1 of the English alphabet would be ABCDEFGHIJKL, 12-mer 2 would be BCDEFGHIJKLM, and the third 12-mer would be CDEFGHIJKLMN, and so on.[12] The Filter develops a growing database of k-mers, and each new k-mer is assessed for its presence (novelty score of 0, i.e., already in the database, therefore: not new) or absence (novelty score of 1, i.e., not in the database, therefore: new) in the

---

[9]Indeed, as Alberto Piazza reflects in his Afterward to Franco Moretti's *Graphs, Maps, and Trees,* literary theorists and biologists share a great many metaphors: Scientists and philosophers alike rely on the alphabetic sequence and, even, the metaphor of poetic inspiration to describe complex, evolutionary problems; DNA is translated and transcribed, not to mention written, copied, and, increasingly, edited. See Franco Moretti, *Graphs, Maps, and Trees.* (New York: Verso, 2007).

[10]For more on the ways that Bloom Filter has been used to measure genetic novelty, see Jason Pell, Arend Hintze, Rosangela Canino-Koning, Adina Howe, James M. Tiedje, and C. Titus Brown, "Scaling Metagenome Sequence Assembly with Probabilistic de Bruijn Graphs." *Proceedings of the National Academy of Sciences* 109, no. 33 (2012): 13272-13277.

[11]The Bloom Filter and the models that we employ in this essay are similar to the sorts of sequence analysis that has recently found its way into computational studies of culture. See Alexander T.J. Barron, Jenny Huang, Rebecca L. Spang, and Simon DeDeo, "Individuals, Institutions, and Innovation in the Debates of the French Revolution." *Proceedings of the National Academy of Sciences of the United States of America* (2018); Shaobin Xu, David A. Smith, Abigail Mullen, and Ryan Cordell, "Detecting and Evaluating Local Text Reuse in Social Networks." *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes, Association for Computational Linguistics* (2014): 50-57; and Richard Jean So, Hoyt Long, and Yuancheng Zhu, "The Dark Code: Race, Computation and the History of White-Black Literary Relations, 1880-2000." *CA,* forthcoming.

[12]Punctuation and spaces are included in each k-mer, which are encoded according to their ASCII value using a modulo-32 operation to reduce the number of bits required to encode each character from 8 to 5.

database. The Bloom Filter is a data structure that achieves a workable efficiency at the cost of the occasional "hash collision," when the filter stores two pieces of information in the same spot, indistinguishably. So, the Bloom Filter will return false positives (identifying an unknown sequence as known), but never false negatives (identifying a known sequence as unknown). That is to say, while the Bloom Filter might erroneously mislabel a segment of text as "not novel," it never erroneously labels a text as "novel." Our novelty scores, therefore, are quite trustworthy, if erring on the conservative side. A text will remain 100% novel until the point at which a k-mer is repeated at any point in the text.[13]

Consider this scan of Gertrude Stein's "A rose is a rose is a rose." (Table 1.) For ease of illustration, spaces are represented as underscores. While segments of Stein's sentence repeat - "rose"- only a perfect match of a k-mer is scored a 0.

A_rose_is_a_	Novel	1
_rose_is_a_r	Novel	1
rose_is_a_ro	Novel	1
ose_is_a_ros	Novel	1
se_is_a_rose	Novel	1
e_is_a_rose_	Novel	1
_is_a_rose_	Novel	1
is_a_rose_i	Novel	1
s_a_rose_is	Novel	1
_a_rose_is_	Novel	1
a_rose_is	Not novel	0
_rose_is_a_r	Not novel	0
rose_is_a_ro	Not novel	0
ose_is_a_ros	Not novel	0
se_is_a_rose	Not novel	0
e_is_a_rose.	Novel	1

Table 1. Novelty Scores

This example is illustrative of the importance of character sequence and order to the Bloom Filter. Though Stein repeats only three words, the word-sequence and punctuation are vital to the Filter's understanding of the text's relative novelty. If more variation were to enter this sequence, the filter would register novelty at the point in which the variable character was introduced.

As this example demonstrates, approximately 59% of Stein's sentence is novel, at the level of character sequence. But imagine this scan occurring at the scale of the chapter, or the text. While variations of "a rose is" may not appear often, certainly character names ("Alice_B_To"), dialogue tags ("_she_said."), or Stein's characteristic repetition would cause her writing to register as less novel as the text progresses. To a Stein scholar, this finding may seem counter-intuitive; Stein's repetition is precisely what makes her work so innovative. However, because the Bloom Filter's assessment of character-sequence novelty ultimately registers

---

[13] All code is available for use in our GitHub Repository: https://github.com/devinhiggins/Novelty-Filter

patterns of relative newness and not-newness over the course of a text, atypical repetition like Stein's can be reclaimed as innovation in its own right, bringing a level of nuance and sophistication to our understanding of novelty beyond a simple proportion of unique words. By registering novelty at the level of the k-mer and the character, the Bloom Filter attends more closely to the stylistic components of text that are often elided, dismissed pragmatically as noise, reading (as in DNA) the most elemental components of language as raw data, from which we might make claims at the level of sentence, the paragraph, the chapter, the text. As a heuristic, we believe that the Bloom Filter's analysis of the smallest, most basic units of measurement in language helps us to peer in at something decidedly less basic, and, potentially impossible to ever fully capture in an algorithmic method: the sense of newness.

Results: Visualizing Intratextual Novelty

What does the Bloom Filter reveal at the scale of the book? What formal patterns emerge in innovation and repetition? And how do those patterns help us understand intratextual novelty? In order to process even larger pieces of data, the sequences of 0s and 1s is binned into intervals of 10,000 k-mers, computing the fraction of 1s in each interval as its novelty score. These scores can then be graphed, providing us with a general shape of how intratextual novelty unfolds in narrative time, from cover to cover. Each text begins with 100% novelty - none of it has been seen before - and novelty gradually decays as character sequences are repeated.[14] The Bloom Filter reveals the ways in which character-sequence novelty unfolds throughout a text, how quickly and often k-mers are reused. The Bloom Filter thus registers moments of relative novelty, in which moments of newness are measured against the aggregated text that came before - or, novelty in context.

---

[14]While it might be theoretically possible for novelty to steadily increase over a text, it is highly implausible; every graph produced by the Bloom Filter has a negative slope.
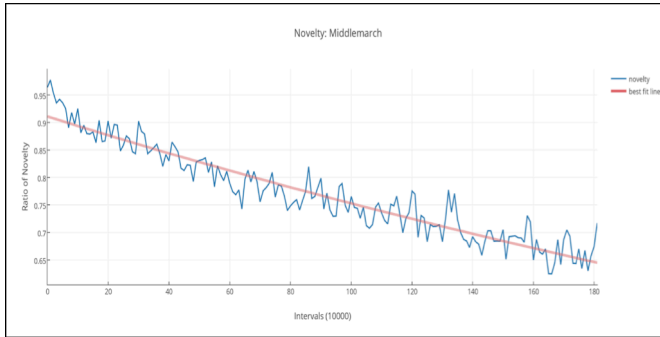
Figure 1. This graph shows the course of novelty in George Eliot's *Middlemarch*. Along the x-axis are intervals of binned k-mers; the y-axis shows the degree of novelty. We see a fairly steady rate of decay, without much internal variation.

In *Middlemarch*, for instance, novelty appears to decay at a steady, gradual rate from the novel's early intervals (see fig. 1). The peaks registering "novel" and valleys registering "not-novel" do not vary substantially from the best fit line (r^2=0.8903), and occur at what appear to be fairly well-paced intervals. This suggests that the language of *Middlemarch* is fairly consistent, and exhibits very little internal variation. *Middlemarch* demonstrates usefully the limitations of the Bloom Filter. It would be a mistake to read this graph as a visual description of the shape of the novel's plot, sentiments, or narrative structure. Though novelty seems to increase in the novel's last, small interval, we cannot rightly say - either by interpreting this graph or by reading *Middlemarch* - that the novel has a surprise ending. We can say, however, that the language in *Middlemarch* appears to vary very little, and novelty appears to decay at a steady rate.

By contrast, and in keeping with our filter's name, we turn to James Joyce's *Ulysses*. How would the filter register *Ulysses*'s novelty, if at all? Bloomian resonances aside, *Ulysses* is a useful counterpoint to *Middlemarch*: it's a common touchstone in discussions of novelty and the signal text in literary modernism. Levenson uses *Ulysses*' "telegraphic repetition" to make his case for modernist intratextuality. How would the novelty filter respond to a text that we presume, from the outset and without question, to be incredibly novel in distinctly modernist terms? Joyce's language is intentionally playful and experimental, providing ample opportunity to test the Bloom Filter's precision and flexibility for measuring intratextual novelty: *Ulysses* relies on a new language system for (almost) each episode, from the advert-speak of "Aeolus" to the catechism of "Ithaca." How does the Bloom Filter respond to these changes?
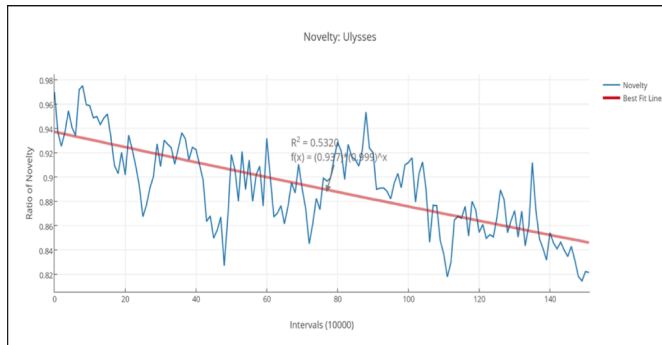
Figure 2. This figure shows the course of novelty in James Joyce's *Ulysses*. We see a high level of variation in the novel's language, as the novelty scores do not conform to the best fit line.

Unlike *Middlemarch*, the dispersal of character-sequence novelty throughout *Ulysses* appears uneven (see fig. 2). While the novelty of *Middlemarch* decays at a steady rate, *Ulysses* appears to exhibit a good deal of variation, suggesting both frequent innovation and repetition. While the graph of *Ulysses* is striking in its inconsistency, we need to return to the text in order to study these high and low points in detail. We wrote a Python script that marks up texts with intervals and running novelty scores to ease the process of cross-referencing and determining what textual features may have produced a given score (available for public access in our GitHub Repository).[15] An examination of *Ulysses* helps us to clarify the nature of the novelty laid out by the Bloom Filter: the percentage of the text that registers as "novel" is far less significant to our purposes than the ways that the text varies - how it repeats, revises, and recirculates textual material.

After the early chapters (declining as Stephen Dedalus and Buck Mulligan speak, then jumping up again as the narrative shifts to Eccles Street), the major spike in novelty occurs around interval 90, approximately 55% of the way through the text. Even given Joyce's experimentation, this increase in novelty is dramatic— particularly at the novel's midpoint. Interval 90 corresponds to episode 14, "Oxen of the Sun." An episode known for its difficulty, "Oxen of the Sun" rehearses the history of the English language as Bloom waits at the Holles Street maternity hospital during Mrs. Purefoy's labor and delivery. Yet, at some point in episode 14, the English language catches up to Joyce, and novelty begins to decline once more.

---

[15] See script and marked up texts at https://github.com/devinhiggins/Novelty-Filter/blob/master/texts/modernism/Joyce_Ulysses_intervals.txt

Around interval 45, *Ulysses*' character-sequence novelty dips to a surprisingly low level. This decline corresponds to the beginning of episode 10, "Wandering Rocks," which presents a series of short vignettes about Dublin residents loosely connected to the Bloom and Dedalus families. The language is highly repetitive, both in word choice and sentence structure, with the character's full name in the subject position - "Father John Conmee" or "Blazes Boylan," without abbreviation - followed by a simple past tense verb (Father John Conmee walked, Father John Conmee thought, Father John Conmee smiled). The repetition in names and sentence structure would seem to account for this apparent drop in novelty. While "Wandering Rocks" demonstrates very little innovation in the form of (say) portmanteau or onomatopoeia, its character-sequence repetition in simple sentences represents a marked difference within *Ulysses*, particularly on the heels of the absurd rhetorical battle of episode 9, "Scylla and Charybdis." These low-points are significant, and should not be read as "not novel," but, as in our example of Gertrude Stein, moments of unusual character-sequence repetition that serve as a form of experimentation itself.

Just as peaks in novelty scores indicate moments of high innovation, we suggest reading these valleys of moments of experimental repetition, a form of textual play and reinvention. The Bloom Filter most effectively highlights the interplay of sameness and difference that is a hallmark of intratextuality. In the context of fluctuating difference, sameness or repetition rightly appears as a formal choice. Because the Bloom Filter produces a measurement of *relative* novelty, moments of intense repetition (such as "Wandering Rocks") only make the bursts of new language (such as "Oxen of the Sun") all the more startling, and the spikes in our readouts steeper. The Bloom Filter thus helps us to positively value unusual repetition as a form of novelty in itself.

It is tempting to read the novel's plot into the graph - to equate each peak with a new episode in Leopold Bloom's walk around Dublin. This simply is not the case; there is no direct correlation between novelty score and plot. While there are moments where the novelty scores correspond to important episodes in the novel, this is due to the language that Joyce employs in a particular episode; this clear correspondence would likely not appear in texts that are more singularly voiced. (Later examples will further illuminate this principle). Even still, the case of *Ulysses* helps us to clarify the type of novelty that the Bloom Filter can register at the scale of the single text. It would seem that it is especially well-suited to evaluating intratextual novelty in highly fragmentary, multivocal texts, like *Ulysses* - evaluating the ways that a text deploys novelty internally through reinvention, regeneration, or recycling textual material.

Results: Comparing Intratextual Novelty

Our first-pass, visual comparison between *Middlemarch* and *Ulysses* appears to indicate that *Middlemarch*ian novelty decays relatively consistently throughout and *Ulysses* displays wild intratextual variation. But how consistent or wild are each of these texts when compared to others? To understand the relative intratextual novelty of both texts, we compared intratextual novelty scores on a mid-sized corpus. We rely on two primary metrics (depicted in the linear regressions above) to compare the intratextual novelty of multiple texts: r^2 and slope. Taken together, these two measures are surprisingly descriptive.

Degree of Formal Variation: We use r^2 to describe the degree of formal variation. The r^2 value, typically a measurement of error, measures the relationship between the best-fit line and the running novelty scores. A high r^2 value (close to 1) indicates very little internal variation since the the best-fit line adequately describes the general shape of the novelty's decay (as in *Middlemarch*). A low r^2 value indicates that there is a high degree of internal variation, for the best fit line does not accurately describe the running novelty scores (as in *Ulysses*). The r^2 can handily describe the degree of variation in the language, accounting for both moments of striking newness and moments of unusual repetition.

Rate of Novelty Decay: By contrast, the slope tells us about the rate or pace of novelty decay, or how quickly a text becomes ingrained in its own linguistic milieu. A steep slope suggests that novelty depletes itself quickly and steadily; the novel's language is likely conventional (unto itself) and consistent throughout—again, *Middlemarch*. A shallow slope, by contrast, indicates that a text decays slowly; said another way, a shallow slope indicates that variation occurs throughout the *entirety* of the text. The slope thus tells us something about the structure of intratextual novelty in the most general terms, without accounting for the shape of its unfolding, as the r^2 value does.

By graphing the r^2 (degree of variation) and slope (rate of decay) together, we can see *how much* and *how often* variation occurs over the course of a text, allowing us to better clarify the contours of intratextual novelty. These two parameters provide a basis for comparing texts to one another (i.e., the intratextual novelty of Text A against the intratextual novelty of Text B) and have the potential to enable macro-level questions. By graphing the degree of variation against the rate of decay, we hypothesized, we could establish a set of parameters for typical intratextual novelty—that is, the way that *most* texts exhibit novelty internally—while accounting for variation in textual patterns.

We ran the Bloom Filter against a corpus of texts, graphing each according to

its degree of variation (x-axis) and rate of decay (y-axis). For this exploratory analysis, our corpus remained relatively small, only 410 Anglophone texts, spanning 1700-2016. We used English-language texts provided by the txtLAB 450 corpus.[16] Additionally, we developed a working list of 20th century titles by compiling publicly available doctoral field exam lists from leading universities, arriving at a relatively stable list of widely-held canonical titles, evenly dispersed over time. (See Appendix for a full list.)
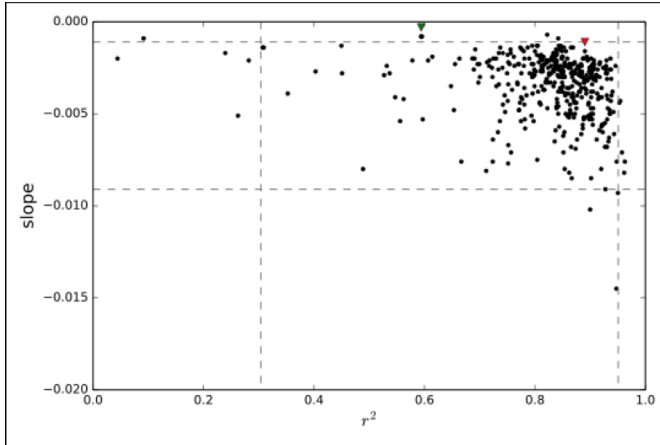
Figure 3. This graph shows the novelty of 410 Anglophone novels. r^2 is measured along the x-axis, and slope is measured along the y-axis. Confidence intervals have been inserted, with p<0.05. *Ulysses* is marked with a green arrow, and *Middlemarch* is marked with a red arrow.

On the x-axis is the degree of variation (r^2), and on the y-axis, the rate of decay (slope). Texts that are closer to 1.0 on the x-axis exhibit a smaller degree of internal variation. The closer a text is to 0.00 on the y-axis, the slower the rate of decay. Theoretically, the text with the highest levels of intratextual novelty would be one with a completely flat slope and an r^2 value of 1.0 - high degrees of variation, consistent throughout the text. A more realistic version of high intratextual novelty, however, seems to couple very flat, slightly negative, slopes with correspondingly high degrees of internal variation. In figure 3, dotted lines represent confidence intervals of $p < 0.05$; texts that fall outside of those dotted lines exhibit statistically significant levels of intratextual novelty. The centermost segment of the graph, where most texts fall, indicates a fairly standard degree of intratextual novelty:

---

[16] Our thanks to Andrew Piper and the McGill University .txtlab for both sharing their post-1923 corpus, and making the Novel 450 corpus available for further research.

a moderate degree of internal variation and a moderate rate of decay. *Middle-march* (marked with the red arrow) is positioned within this central cluster, and is typical of novels in our corpus.

Many of the texts that exhibit an unusual degree of intratextual novelty have long been held up as exemplars of modernist novelty, giving us confidence in the validity of our measure. Yet many of these results surprise us. *Ulysses*, marked by a green arrow, is not quite as overwhelmingly novel as we presumed; while exhibiting a slower rate of decay, its degree of internal variation is actually not *statistically* significant to the degree we're measuring here. This is not to say that *Ulysses* is not a significantly novel text according to the Bloom Filter, but that different types of intratextual novelty emerged through our analysis.

1. Intratextual Fragmentation: High, Sustained Variation

Novels that have both a significantly slow rate of decay and a significantly high degree of internal variation are our most novel texts, and are located in the top left corner of our graph; tellingly, only one novel falls in this region: *The Sound and the Fury*. No other text in our corpus exhibits anything close to this degree of intratextual novelty, in terms of both variation and continuity. (In a larger corpus, we believe that it is entirely plausible that other texts would exhibit novelty of this degree.)
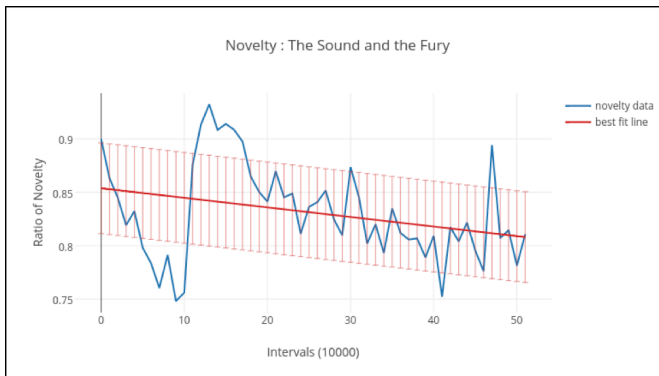

Figure 4. This graph shows the course of novelty over William Faulkner's *The Sound and the Fury*. Within our corpus, The Sound and the Fury is by all measures novel, with a significantly shallow slope and a significantly low r^2. The bars on this graph show how often novelty falls outside standard margins of error.

No other text in our corpus contains an interval score that exceeds that of the first interval; novelty consistently decays from the beginning of most texts. Here, however, the novel begins with a novelty score of .8998 in interval 1, then drops to a low-point of .748 at interval 9, before jumping up to .932 at interval 13. An examination of the binned text confirmed our suspicions: *The Sound and the Fury* famously opens with narrative written from the perspective of Benjy Compson, which consists of a great deal of word-level repetition, a limited vocabulary, and simple syntax. The novelty score begins shifting upward at interval 10, corresponding with the shift to Quentin Compson's point of view; novelty increases to a 1.0, entirely new in comparison to Benjy's narration, reflecting not only the point of view shift but the difference in language and syntactic complexity. While this result is consistent with a reading of the novel, and is a (more extreme) version of what we expected to see reflected in a graph of *The Sound and the Fury,* it is less the individual high- and low-points that are significant than the *degree* of difference between the two. To say that Benjy's point of view is somehow not-novel because the scores are so low is simply false; indeed, Benjy's narration is perhaps the most unusual aspect of *The Sound and the Fury*. Similarly, to say that Quentin's point of view is somehow more novel simply because of its high score is also disingenuous. The difference between these two sections, and the rather jarring shift, is what distinguishes *The Sound and the Fury* from other novels in our corpus.

The Sound and the Fury is also continuously novel, spiking again at interval 47; that is to say, the novel repeatedly reinvents itself, introducing new types of textual variation relatively late in the text. While the middle sections of the novel, narrated by Quentin and Jason Compson, do not vary substantially (all scores falling within the standard error range), the end of the novel peaks toward the beginning of "April Eighth, 1928." This may be due to the narrative emphasis on Dilsey and the prevalence of Faulkner's variation on black English; it may also be due to the shift from first person to third. Because the language shifts with each narrative perspective, and because the novel changes perspective continuously, *The Sound and the Fury* exhibits an especially flat slope. More than any other text we examined, *The Sound and the Fury* exhibits high and consistent textual variation, whereas other novels with multiple narrators (as in our next case, *The Color Purple*) taper off once the diverging voices have been introduced.

That *The Sound and the Fury* should emerge as our most novel text confirmed our confidence in our measure's ability to capture the types of intratextual novelty most closely associated with modernism - novelty akin to fragmentation, montage, or a composite form. *The Sound and the Fury* is a fascinating example; no other text in our corpus match its internal variation or its consistent renova-

tion. But it is also important to note that *The Sound and the Fury* is not the only Faulkner text in our corpus. We also included *As I Lay Dying,* for comparison's sake, and found that it falls within the typical range. Our intratextual novelty scores seem to be not necessarily a matter of oeuvre or aesthetic, but (appropriately) vary by text. For now, *The Sound and the Fury* is in a class of its own, a text that varies wildly in its Bloomian novelty, but when all is said and done, the final page turned, has also left a record of sustained textual invention.

2. Dialogic Novelty: High Variation, Average Decay

Novels in Category 2 exhibit many of the same features as *The Sound and the Fury* by virtue of their equally high degree of internal variation: the language varies such that the best fit line does not accurately describe the majority of the text. While more texts exhibit this type of novelty than Category 1 novelty, it is still rather rare in our corpus. The texts that fall into this category, including *The Color Purple* and *A Portrait of the Artist as a Young Man*, would indicate that our results fall less neatly into categories than a continuum. We turn to *The Color Purple* for illustration. While the significance of *A Portrait of the Artist* is unsurprising, given our research team's investment in literary modernism and our measure's apparent sensitivity to intratextual fragmentation, *The Color Purple* is associated neither with the modernist period, nor is it held up as an example of a modernist continuation beyond midcentury into the latter 20th. A closer investigation of Walker's novel is particularly illustrative of the Bloom Filter's limitations.

The Color Purple has a typical rate of decay. But the degree of internal variation is significantly high. *The Color Purple* is an epistolary novel, and like *The Sound and the Fury*, has more than one first-person narrator—Celie composes the majority of letters, first to God and then to her sister Nettie, who in return writes a number of letters to Celie. There is a fairly steep rate of decay in the first quarter of the novel, consistent with Celie's vocabulary and short letters. The novelty increases, however, once Celie discovers and begins reading Nettie's letters, hidden over a number of years (the biggest spike in novelty corresponds with Celie's *initial* discovery). Nettie's letters employ a more sophisticated vocabulary and sentence structure, in keeping with her education. Furthermore, Nettie's vocabulary is contextually different than Celie's. Nettie is a missionary in northern Africa, and her vocabulary reflects her immediate cultural context; Nettie's letters introduce nouns that would never occur in Celie's letters. Even still, while the running novelty score peaks slightly each time the narration shifts between the sisters, it

recalibrates to their respective vocabularies and contexts, decaying at a standard rate.
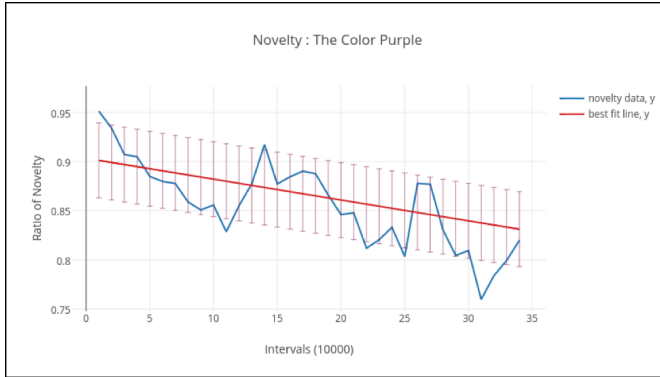
Figure 5. This graph shows the novelty of Alice Walker's *The Color Purple*, which exhibits a significantly low r² but a typical slope.

The instances of unusual repetition in *The Color Purple* are especially instructive—the name "dialogic novelty" is meant to gesture toward these low-points as much as the novel's structural dialogue between Celie and Nettie. Unlike Nettie's letters, Celie's, as the main narrative, consist of a great deal of dialogue. Celie writes without punctuation and with a great deal of repetition to indicate who is speaking and to whom. This passage, for example, comes from the interval with the lowest novelty score:

> Too late to cry, Miss Eleanor Jane, say Sofia. […]
>
> You just don't like him cause he look like daddy, say Miss Eleanor Jane.
>
> You don't like him cause he look like daddy, say Sofia. I don't feel nothing about him at all. I don't love him, I don't hate him. I just wish he couldn't run loose all the time messing up folks stuff.
>
> All the time! All the time! say Miss Eleanor Jane. Sofia, he just a baby. Not even a year old. He only been here five or six times.
>
> I feel like he been here forever, say Sofia.

This back and forth between Sofia and Eleanor Jane is fairly typical of the dialogue that Celie writes. With a sort of call-and-response structure, Sofia repeats much of what Eleanor Jane tells her, with slight differences, often implied tonally ("You just don't like him cause he look like daddy," and "You don't like him cause he

look like Daddy"). Without punctuation, Celie also insists on repetitive dialogue tags—"say Sofia" and "say Miss Eleanor Jane" repeat a number of times in this passage, consistent with all of the dialogue in the novel. Here, the measurement is somewhat deceptive. Dialogue is not the same thing as experimental repetition, in a Steinian sense, but the Bloom Filter, agnostic to the fictional content it scans, encodes these segments of texts similarly. This is not to say that *The Color Purple* is a false positive, but it does reveal the limitations of the measurement. The work of close reading remains necessary in order to understand the dynamics being measured, scored, and graphed.

3. Maximalist Novelty: Average Variation, Slow Decay

Category 3 texts demonstrate a statistically significant level of novelty across one variable: slope. Category 3 texts sustain a novelty continuously over their length, but do not exhibit as much internal variation as the unusual exemplar in Category 1; interval novelty scores rarely exceed or fall below a standard margin of error, resulting in an non-significant r^2 score. The slopes, however, are statistically flatter than most texts in our corpus. Novels in this category include *Ulysses*, *Infinite Jest,* and *Gravity's Rainbow*. We'll take *Infinite Jest* as our example.

Infinite Jest (see Fig. 6) exhibits very little internal variation; its r^2 value falls within the typical range. Initially, the novel's typical r^2 score surprised us. We were expecting *Infinite Jest* to exhibit a higher degree of internal variation. In contrast to *The Sound and the Fury* and *Ulysses* (which exhibits a low, though still statistically insignificant r^2) *Infinite Jest* is more univocal on a narrative scale; while *The Sound and the Fury* and *The Color Purple* are focalized in the first-person and through a number of characters, *Infinite Jest* retains a consistently detached third-person narration. *Infinite Jest* owes its flat slope and significant score to one distinctive portion of the novel: the end. Those familiar with the novel can likely hazard a guess why: the novel concludes with a dense and meticulously detailed appendix, beginning around interval 284. The "Notes and Errata" novelty scores are, true to their name, much more erratic.
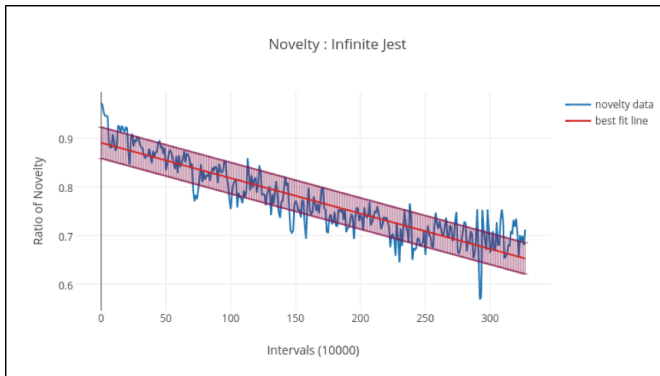
Figure 6. This graph shows novelty over the course of David Foster Wallace's *Infinite Jest*. While the r² is typical, the slope is significantly shallower than most texts in our corpus.

The endnotes have dramatically influenced the shape of the novel, pulling the slope upward with their strange variations in form. Other texts that fall into this category also reflect this pattern, with the later portions of the text shifting in novelty and rendering the slope more shallow. Thomas Pynchon's *Gravity's Rainbow* also sits comfortably in this category. *Gravity's Rainbow* (see Fig. 7) begins with a rather steep decline. Around interval 125, however, the novelty levels off; a best-fit line drawn from interval 125 to the end of the novel would look relatively flat, easily characterizing the majority of the novel. While perhaps Pynchon and Wallace do not exhibit a great deal of innovation on the micro-level - or, at least, not much more than is typical - what novelty they do exhibit comes later in the text. *Infinite Jest*'s appendix is a limit case; we hypothesize that additional texts in this category would likely resemble something closer to *Gravity's Rainbow* than *Infinite Jest*.
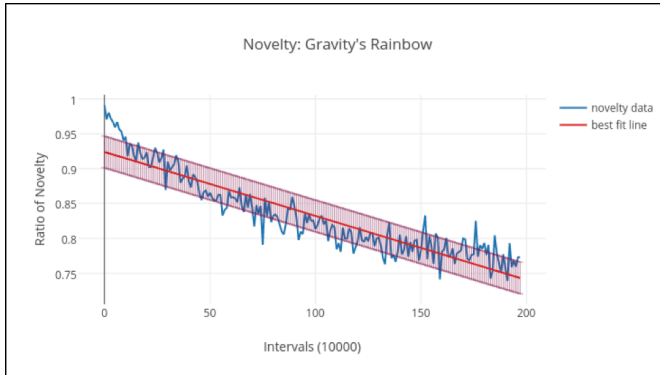
Figure 7. This graph shows novelty over the course of Thomas Pynchon's *Gravity's Rainbow*. While the r² is typical, the slope is significantly shallower than the most texts in our corpus.

Our examples in this category raise the question: is this *postmodern* novelty? Does it speak to characteristics of the postmodern novel, typified by Wallace and Pynchon? While the presence of Ulysses seems to confound a neat period-specific reading, it perhaps suggests a continuity of literary experimentation in the twentieth century. As exemplars of their respective periods, the similarity of these texts may gesture towards a more singular modernity than a postmodern rupture. Rather than a question of periodization, we might also approach these texts with questions of form. These three texts are all quite long; perhaps this type of novelty has more to do with length than with any other stylistic category.[17] Perhaps we might say, instead, that the novels in this category, *Ulysses* included, represent a sort of maximalist tendency that reached its apex in postmodernism.[18]

---

[17] One would be forgiven for wondering at this point whether the Bloom Filter is simply biased toward longer texts: A line drawn from the novelty reading at point A (the beginning of the text) to point B (the ending) has a shallower slope the farther apart they are. Yet our data reveals that the longer texts we examine here really are generating as much new textual novelty over their length than their shorter yet otherwise identical counterparts. A bias in favor of "more novel" longer texts would be operative only if the novelty decay plateaued at some point, in which case extending the plateau a few hundred more pages would certainly seem to "artificially" elevate the calculated slope. Our analysis shows, however, that attempts to fit non-linear patterns of decay to the data do not return increased accuracy, and we do not see evidence of this plateauing effect. In fact, the high r² values our maximalist texts tend to receive, indicate that their shallowly sloping novelty does not subside in ways a non-linear measure would capture.

[18] Stefano Ercolino has remarked on the relationship between length and maximalism, arguing that, "Length is not simply a neutral material aspect as regards the maximalist novel, but something more… It is a possibility that turns out to be related both to the strongly innovative and experimental nature of maximalist novels and to their ambition to realize synthetic-totalizing representations of the world." See Ercolino, "The Maximalist Novel." *Comparative Literature* 64.3 (2012): 241-256.

So, just as the Bloom Filter is particularly well-suited to uncovering patterns in fragmentary texts, it may also be responsive to the long novels that characterize postmodernism. If not "postmodern" and if not simply "long," then perhaps we can see in this type a definition of Maximalist Novelty.

4. Not Novelty: Very Little Variation, Fast Decay

Curiouser and curiouser. *Alice's Adventures in Wonderland* (see Fig. 8) provides us with another fascinating example. Like *The Sound and the Fury*, *Alice's Adventures in Wonderland* is the only text of its type, though others (including Carroll's *Through the Looking Glass*) do tend toward this type of novelty. While *The Sound and the Fury* has a uniquely low r^2 and shallow slope, *Alice's Adventures* is nearly its inverse: it has the steepest slope of any text in our corpus, though the r^2 is quite high; at times, there is very little variation between the best fit line and the novelty interval scores. This suggests that there is very little linguistic variation at the structural or vocabulary level.

Alice in Wonderland is not significant because of its high degree of internal variation or its slow decay; it is unique in terms of how little variation it exhibits. This may be related to any number of factors: its genre, its target audience (it is the only children's book in our corpus), its precocious narrational style. While these hypotheses are provocative, and may bear out with more testing, none of them seem to quite explain why *Alice* scores so distinctly according to the Bloom Filter. *Alice's Adventures in Wonderland* offers an important corrective: all texts exhibit some novelty, according to our measure. To argue that the texts that fall within the typical novelty range are not novel is incorrect; as measure of intra-textual novelty, the Bloom Filter gives each text an unfolding series of scores for how novel it is according to its own parameters. *Alice's Adventures in Wonderland* registers as significant for how little novelty it exhibits in comparison to our range of "normal" texts.
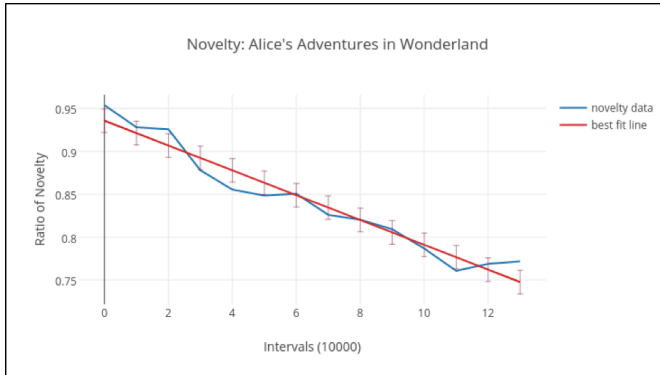
Figure 8. This figure shows the novelty of Lewis Carrol's *Alice's Adventures in Wonderland*. This text is significant for its relative lack of novelty: the slope is quite steep, and the r²is quite high.

While only those texts that fall outside of the dotted lines are significantly novel, it is worth considering this typology as poles toward which novels tend on a continuum: novels with a lower *r² tend* to vary more syntactically; novels with a steeper slope *tend* to be more conventional. Though tendencies in novelty may not be statistically significant, they are nevertheless descriptive as we seek to understand the shapes that intratextual novelty takes, and provide a useful basis for comparing novelty across texts. By no means is this analysis exhaustive; rather, it offers an explanatory heuristic by which we might consider some basic shapes, patterns, and behaviors of intratextual novelty. Without question, there are more questions to be asked regarding languages, genres, and any number of other analytical categories.

Conclusion

This proof-of-concept has demonstrated the potential of the Bloom Filter to measure intratextual novelty and suggested ways that the results of this measurement might be brought to bear on larger questions of literary history, helping us to better understand the paradoxical concept of novelty and the literature that lays claims to it. Yet our results raise far more questions than they answer, first among them: does novelty really exist at the level of the alphabetic character? William James insisted that not even *meaning* could be drawn from the character level:

"It is not as if men had first invented letters and made syllables out of them, then made words of the syllables and sentences of the words; - they actually followed the reverse order."[19] Yet words are also not made out of thin air - they do take form in specific and concrete languages obeying physical rules of morphology and phonology that do leave their traces in the accidents of spelling.   Still, can we argue that novelty can be measured at the level of the k-mer, an admittedly constructed unit of measurement? Certainly, the k-mer is not the "it" that Ezra Pound had in mind when he insisted upon making it new, not Eliot's "really new" that should disrupt and rearrange the nature of tradition. And, certainly, novelty cannot be disentangled from its cultural context. This measurement is an imperfect one, without question. Imperfect though the measurement might be, we believe that the Bloom Filter does provide a suitable proxy for the measurement of intratextual novelty, at both the micro and macro scales.

Beyond demonstrating the effectiveness of Bloom Filter, our preliminary analysis raises a number questions, inviting further application of the measurement that can intervene in the literary history and periodicity of the 20th century. Consider two of our significantly novel texts: *Infinite Jest* and *Gravity's Rainbow*; intratextual novelty may, in fact, be more characteristic of postwar novels, prompting any number of questions regarding postmodernism's extension, negation, or response to earlier modernist experimentation. While, on the one hand, these results may affirm a lineage of 20th century maximalist experimentation—from Joyce to Pynchon to Wallace—they also seem to disrupt the periodizing narrative of modernism's conclusion on or about 1945.

Alternatively, it seems equally plausible that, when taken as a whole, modernism may be far less novel than we thought—that these statistically significant texts are, in fact, outliers. Recently, scholars of literary modernism have moved away from novelty as an explanatory feature of periodicity, instead considering modernism's relationship to the mundane, the quotidian, the obsolete, the ordinary, and the everyday.[20] No connection between modernism and intratextual novelty would, in itself, be a fascinating result. While modernism is often defined by such figures as Joyce and Faulkner, a larger corpus of modernist fiction could reveal the exemplarity or exceptionality of their writing in the context of their

---

[19]William James, "The Compounding of Consciousness" in *Writings: 1902-1910* (New York: Library of America, 1987): 715.

[20]Mark Goble writes, "It might even be the case that modernism persists for us today largely as an aesthetic of obsolescence… precisely because we no longer feel required to insist on modernism's novelty as the most important measure of its value," (147). See "Obsolescence" in Eric Hayot and Rebecca Walkowitz, *A New Vocabulary for Global Modernism.* (New York: Columbia UP, 2016): 146. See also Liesl Olson, *Modernism and the Ordinary.* (New York: Oxford UP, 2009). Juan A Suárez, *Pop Modernism: Noise and the Reinvention of the Everyday.* (Champagne: U Illinois P, 2007). Bryony Randall, *Modernism, Daily Time and Everyday Life.* (New York: Cambridge UP, 2007).

contemporaries, widening or narrowing the Great Divide accordingly.

Finally, what of *The Color Purple*? Among our statistically significant results, it is the only novel written by a woman, and the only novel written by a person of color—authors who tend to be written out of the history of High Modernism (capital-H, capital-M) and, for the most part, postmodernism. On the one hand, *The Color Purple*'s presence may suggest limitations of the measurement—dialogue mistaken for formal innovation. But perhaps the significance of *The Color Purple* prompts a reconsideration of the whitened and masculinized genealogy of 20th century experimentation. Perhaps Walker's formal innovations, as revealed by the Bloom Filter, challenge notions of literary experimentation and novelty.

While these questions are provocative, they cannot be answered without expansion of this study and more creative applications of the Bloom Filter. As with any new form of measurement or hypothesis, only replication will validate the hypotheses we propose here. But we believe that this method is promising, particularly for scholars of literary modernism, and it is our hope that even more novel uses for the Bloom Filter might emerge in time.

Appendix: Corpus

PubYear	AuthorName	Title	intervals	lexicaldiversity	r2	slope
1958	Achebe,Chinua	ThingsFallApart	28	0.103	0.8736	-0.0069
1972	Adams,Richard	WatershipDown	92	0.061	0.7903	-0.0028
1907	Ade,George	TheSlimPrincess	12	0.202	0.8533	-0.006
2008	Adiga,Aravind	TheWhiteTiger	46	0.103	0.7345	-0.0034
1847	Aguilar,Grace	HomeInfluence	96	0.055	0.8134	-0.0026
1850	Aguilar,Grace	ValeofCedars	54	0.097	0.848	-0.0036
1872	Ainsworth,William	TheRoyalOak	96	0.081	0.833	-0.0026
2012	Albom,Mitch	TimeKeeper	21	0.153	0.9478	-0.0076
1869	Alcott,Louisa	LittleWomen	100	0.061	0.8324	-0.0021
2010	Alexie,Sherman	WarDances	24	0.153	0.6533	-0.0048
1868	Alger, Horatio	RaggedDick	26	0.1	0.7974	-0.0054
1879	Alger, Horatio	TelegraphBoy	20	0.103	0.7365	-0.0054
2013	AlireSaenz,Benjamin	EverythingBeginsandEndsattheKentucky	32	0.089	0.8372	-0.0059
1864	Anon	FromPillarToPost	33	0.131	0.9375	-0.0026
1940	Armstrong,Margaret	Trelawney	148	0.066	0.2623	-0.0051
1891	Atherton,Gertrude	AQuestionOfTime	21	0.152	0.8102	-0.0045
1900	Atherton,Gertrude	SenatorNorth	56	0.089	0.9115	-0.0029
2013	Atkinson,Kate	LifeAfterLife	81	0.106	0.7972	-0.0028
1985	Atwood,Margaret	TheHandmaidsTale	53	0.102	0.7753	-0.0028
1811	Austen,Jane	SenseandSensibility	67	0.064	0.8192	-0.0033
1813	Austen,Jane	PrideandPrejudice	68	0.057	0.8506	-0.0035
1813	Austen,Jane	PrideandPrejudice	68	0.056	0.8336	-0.0035
1814	Austen,Jane	MansfieldPark	88	0.053	0.8699	-0.0028
1814	Austen,Jane	MansfieldPark	88	0.053	0.8698	-0.0028
1815	Austen,Jane	Emma	88	0.063	0.8675	-0.0027
1916	Baily, Waldron	Thehomewardtrail	33	0.135	0.9158	-0.0041
1878	Ballantyne,RM	InTheTrackOfTheTroops	51	0.113	0.8213	-0.0022
2005	Banville,John	TheSea	33	0.153	0.9318	-0.0042
2011	Barnes,Julian	SenseofAn	23	0.145	0.934	-0.0063
1885	Barr,Amelia	JanVeeder'sWife	35	0.107	0.8405	-0.0044
1900	Barr,Amelia	TheMaidofMaidenLane	44	0.092	0.9127	-0.0039

PubYear	AuthorName	Title	intervals	lexicaldiversity	r2	slope
1911	Barrie,J.M.	PeterPan	25	0.106	0.7937	-0.0051
1786	Beckford,William	Vathek	21	0.174	0.9269	-0.0068
2006	Behrens,Peter	LawofDreams	63	0.089	0.829	-0.0024
1887	Bellamy,Edward	LookingBackward	44	0.105	0.6479	-0.0035
1902	Bellamy,Edward	Eleonora	42	0.121	0.9143	-0.0042
1964	Bellow,Saul	Herzog	77	0.12	0.8867	-0.002
1902	Bennett,Arnold	GrandBabylonHotel	38	0.104	0.8547	-0.004
2005	Bergen,David	TheTimeinBetween	42	0.086	0.8866	-0.0058
1993	Bickmore,Barbara	DistantStar	100	0.072	0.8607	-0.0025
1913	Biggers,EarlDerr	SevenKeystoBaldPate	43	0.136	0.8341	-0.0027
1869	Blackmore,R.D.	LornaDoone	142	0.054	0.8561	-0.0016
1996	Blake,Jennifer	SilverTonguedDevil	60	0.1	0.9034	-0.0034
1796	Bonhote,Elizabeth	BungayCastle	49	0.089	0.8389	-0.0047
1843	Borrow,George	TheBibleinSpain	124	0.073	0.8107	-0.002
1981	Bowie,Donald	CableHarbor	69	0.111	0.917	-0.003
2008	Boyden,Joseph	ThroughBlackSpruce	70	0.068	0.8851	-0.0028
1862	Braddon,Mary	LadyAudley'sSecret	82	0.076	0.841	-0.0029
1864	Braddon,Mary	HenryDunbar	86	0.073	0.7223	-0.0025
1883	Braddon,Mary	TheGoldenCalf	89	0.084	0.839	-0.0021
1862	Braddon,MaryElizabeth	LadyAudleysSecret	82	0.076	0.8381	-0.0025
1848	Bronte,Ann	TheTenantofWildfellHall	93	0.083	0.8499	-0.0026
1847	Bronte,Charlotte	JaneEyre	104	0.091	0.8442	-0.0015
1847	Bronte,Charlotte	JaneEyre	102	0.08	0.8431	-0.0013
1847	Bronte,Emily	WutheringHeights	66	0.1	0.9305	-0.0031
1847	Bronte,Emily	WutheringHeights	64	0.087	0.9458	-0.0024
1890	Broughton,Rhoda	Alas!	83	0.092	0.8363	-0.0023
1798	Brown,CharlesBrockden	Wieland	47	0.091	0.8579	-0.0043
1799	Brown,CharlesBrockden	ArthurMervyn	84	0.068	0.5781	-0.0021
2013	Brown,Dan	Inferno	91	0.101	0.7628	-0.0024
1856	Browning,ElizabethBarrett	AuroraLeigh	48	0.122	0.8034	-0.0018
1994	Bunn,TDavis	RidersofthePaleHorse	49	0.125	0.8485	-0.003
1994	Bunn,TDavis	RidersofthePaleHorse	49	0.105	0.8672	-0.004
1778	Burney,Fanny	Evelina	86	0.062	0.8258	-0.0029
1782	Burney,Fanny	Cecilia	186	0.039	0.737	-0.0014
1796	Burney,Fanny	Camilla	204	0.041	0.805	-0.0013
1917	Burroughs,EdgarRice	TarzanoftheApes	48	0.091	0.8119	-0.0035
1959	Burroughs,William	Naked Lunch	49	0.142	0.7496	-0.0028
1917	Cahan,Abraham	TheRiseofDavidLevinsky	92	0.07	0.8991	-0.0025
1843	Carlyle,Thomas	PastandPresent	62	0.12	0.5318	-0.0024
1865	Carroll,Lewis	Alice'sAdventureinWonderland	14	0.104	0.9473	-0.0145
1871	Carroll,Lewis	ThroughtheLookingGlass.txt	16	0.108	0.8999	-0.0102
1912	Cather,Willa	Alexander'sBridge	14	0.141	0.92	-0.008
2013	Catton,Eleanor	Luminaries	155	0.073	0.8215	-0.0022
2011	Chan,Darcie	TheMillRiverRecluse	66	0.085	0.768	-0.0033
1857	Chatterton,Georgiana	Lifeanditsrealities	28	0.122	0.9227	-0.006
1901	Chesnutt,Mary	MarrowofTradition	50	0.102	0.8532	-0.003
1836	Child,Lydia	Philothea	42	0.103	0.8748	-0.0044
1890	Chopin,Kate	AtFault	33	0.148	0.8695	-0.0041
1899	Chopin,Kate	TheAwakening	27	0.122	0.8661	-0.005
1899	Chopin,Kate	TheAwakening	28	0.121	0.8714	-0.0044
2008	Christensen,Kate	GreatMan	49	0.126	0.85	-0.0038
2011	Clark,MaryHiggins	IllWalkAlone	51	0.083	0.8748	-0.0047
1930	Coates,RobertM	TheOutlawYears	38	0.148	0.8213	-0.0031
2014	Coben,Harlan	MissingYou	62	0.097	0.7807	-0.0036
1852	Collins,Wilkie	Basil	63	0.084	0.8294	-0.0028
1859	Collins,Wilkie	WomaninWhite	135	0.052	0.812	-0.002
1860	Collins,Wilkie	TheWomaninWhite	135	0.052	0.8122	-0.002
1868	Collins,Wilkie	Moonstone	105	0.055	0.8177	-0.0023
2012	Connelly,Michael	TheBlackBox	62	0.088	0.7005	-0.003
1899	Conrad,Joseph	HeartofDarkness	21	0.159	0.9219	-0.0056
1899	Conrad,Joseph	HeartofDarkness	21	0.159	0.922	-0.0056
1907	Conrad,Joseph	SecretAgent	52	0.114	0.8234	-0.0033
1921	Conrad,Joseph	TheShadowLine	21	0.136	0.9035	-0.0053
1826	Cooper,JameFenimore	TheLastoftheMohicans	83	0.068	0.8947	-0.0032
1823	Cooper,JamesFenimore	ThePioneers	95	0.075	0.8625	-0.0023
1853	Craik,Dinah	Agatha'sHusband	78	0.081	0.8994	-0.0025
1856	Craik,Dinah	JohnHalifax	98	0.081	0.9295	-0.0021
1945	Crane,Frances	TheIndigoNecklae	33	0.101	0.9166	-0.0059
1892	Crane,Stephen	MaggieGirlOfTheStreets	13	0.189	0.4886	-0.008
1895	Crane,Stephen	TheRedBadgeofCourage	25	0.138	0.8439	-0.0057
1895	Crane,Stephen	RedBadgeOfCourage	26	0.138	0.8636	-0.0048
1911	Crockett,Samuel	Theladyofthehundreddresses	42	0.124	0.7131	-0.0022
1898	Crockett,SR	TheRedAxe	59	0.091	0.8864	-0.003
1920	Cullum,Ridgwell	TheheartofUnaga	72	0.076	0.9282	-0.0026
1920	Curwood,JamesOliver	TheValleyofSilentMen	40	0.088	0.905	-0.005
1861	Davis,Lydia	LifeintheIronMills	8	0.208	0.9627	-0.0076
1920	Day,Clarence	TheSimianWorld	8	0.234	0.5559	-0.0054
2013	Day,Sylvia	EntwinedwithYou	60	0.092	0.842	-0.0038
1985	DeLillo,DonWhite	Noise	59	0.133	0.8018	-0.0027

PubYear	AuthorName	Title	intervals	lexicaldiversity	r2	slope
1853	Dickens,Charles	BleakHouse	193	0.051	0.8468	-0.0013
1859	Dickens,Charles	ATaleofTwoCities	75	0.079	0.8459	-0.0024
1861	Dickens,Charles	GreatExpectations	99	0.067	0.8347	-0.002
1861	Dickens,Charles	GreatExpectations	99	0.067	0.835	-0.002
1826	Disraeli,Benjamin	VivianGrey	109	0.069	0.7809	-0.002
1837	Disraeli,Benjamin	Venetia	91	0.062	0.7799	-0.0024
1920	DosPassos,John	ThreeSoldiers	73	0.073	0.696	-0.0023
1921	DosPassos,John	ThreeSoldiers	74	0.073	0.7176	-0.0021
1889	Doyle,ArthurConan	TheMysteryoftheCloomber	26	0.135	0.7506	-0.0043
1891	Doyle,ArthurConan	TheDoingsofRafflesHaw	20	0.137	0.9304	-0.0068
1900	Dreiser,Theodore	SisterCarrie	87	0.068	0.7411	-0.0025
1900	Dreiser,Theodore	SisterCarrie	87	0.068	0.7692	-0.0023
1912	Dreiser,Theodore	TheFinancier	109	0.067	0.7517	-0.0022
1915	Dreiser,Theodore	TheGenius	168	0.052	0.7869	-0.0014
1896	Dunbar,PaulLaurence	TheUncalled	27	0.118	0.6974	-0.0033
1987	Dvorkin,David	Budspy	50	0.11	0.9017	-0.0041
1800	Edgeworth,Maria	CastleRackrent	18	0.122	0.7517	-0.0067
1801	Edgeworth,Maria	Belinda	100	0.06	0.824	-0.0025
1806	Edgeworth,Maria	Leonora	36	0.103	0.9008	-0.0052
1860	Eliot,George	TheMillontheFloss	114	0.068	0.831	-0.0019
1861	Eliot,George	SilasMarner	39	0.108	0.8683	-0.0039
1862	Eliot,George	Romola	128	0.075	0.9069	-0.0021
1871	Eliot,George	Middlemarch	178	0.058	0.8904	-0.0016
1876	Eliot,George	DanielDeronda	174	0.058	0.7949	-0.0013
1952	Ellison,Ralph	InvisibleMan	102	0.083	0.839	-0.0022
1988	Fante,John	FullofLife	21	0.154	0.933	-0.0065
1988	Farelly,Peter	OutsideProvidence	25	0.152	0.9118	-0.0064
1929	Faulkner,William	SoundAndTheFury	52	0.086	0.0915	-9e-04
1930	Faulkner,William	AsILayDying	30	0.102	0.8585	-0.0051
1873	Fawcett,Edgar	Purpleandfinelinen	81	0.09	0.8547	-0.0024
2012	Ferguson,Will	419	64	0.125	0.8774	-0.0024
1920	Fitzgerald,FScott	ThisSideofParadise	41	0.138	0.8811	-0.0027
1922	Fitzgerald,FScott	TheBeautifulandtheDamned	72	0.11	0.9354	-0.0026
1925	Fitzgerald,FScott	TheGreatGatsby	26	0.13	0.942	-0.0049
1934	Fitzgerald,FScott	TenderIsTheNight	61	0.112	0.8795	-0.0025
2014	Flanagan,Richard	NarrowRoad	67	0.093	0.7992	-0.0029
1856	Flaubert,Gustave	MadameBovary	65	0.095	0.9285	-0.0028
1993	Flook,Maria	FamilyNight	42	0.138	0.9053	-0.0047
2012	Flynn,Gillian	GoneGirl	86	0.098	0.7247	-0.002
2012	Follett,Ken	WinteroftheWorld	185	0.066	0.8035	-0.0013
1915	Ford,FordMaddox	TheGoodSoldier	41	0.101	0.8972	-0.0044
1924	Ford,FordMaddox	ParadesEnd	112	0.091	0.8491	-0.0019
1915	Ford,FordMadox	TheGoodSoldier	41	0.101	0.8882	-0.0052
1908	Forster,E.M.	ARoomWithaView	37	0.113	0.9205	-0.0036
1910	Forster,E.M.	HowardsEnd	62	0.093	0.9059	-0.0031
1908	Forster,EM	RoomWithAView	37	0.112	0.9099	-0.0032
1910	Forster,EM	HowardsEnd	61	0.093	0.9168	-0.0023
1996	Foster,David	InfiniteJest	328	0.087	0.8222	-7e-04
1797	Foster,HannahWebster	TheCoquette	29	0.101	0.7815	-0.0053
2014	Fowler,KarenJoy	WeareAll	46	0.115	0.8455	-0.0036
1980	Freed,Donald	TheChinaCard	48	0.148	0.8964	-0.0031
1894	Freeman,MaryWilkins	Pembroke	44	0.091	0.7451	-0.004
1821	Galt,John	AnnalsoftheParish	35	0.11	0.842	-0.005
1848	Gaskell,Elizabeth	MaryBarton	87	0.068	0.9286	-0.0024
1848	Gaskell,Elizabeth	MaryBarton	87	0.068	0.9366	-0.002
1854	Gaskell,Elizabeth	NorthandSouth	99	0.067	0.8538	-0.0022
1854	Gaskell,Elizabeth	NorthAndsouth	99	0.066	0.8503	-0.0019
2012	Giffin,Emily	WhereWeBelong	60	0.104	0.9385	-0.004
1891	Gissing,George	NewGrubStreet	102	0.059	0.8537	-0.0023
1891	Gissing,George	NewGrubStreet	102	0.059	0.8586	-0.002
1893	Gissing,George	TheOddWomen	77	0.072	0.9089	-0.003
1893	Gissing,George	OddWomen	77	0.071	0.903	-0.0025
1794	Godwin,William	CalebWilliams	81	0.068	0.8674	-0.003
1979	Gordimer,Nadine	BurgersDaughter	73	0.127	0.8128	-0.0021
1990	Graham,Heather	TheVikingsWoman	70	0.07	0.8492	-0.0033
1893	Grand,Sarah	TheHeavenlyTwins	156	0.054	0.8379	-0.0017
1916	Grey,Zan	TheBorderLegion	55	0.093	0.8705	-0.0032
1870	Griffith,Cecil	ValentineForde	24	0.123	0.9268	-0.0055
2009	Grisham,John	TheAssociate	61	0.1	0.9048	-0.0031
2006	Gruen,Sara	WaterforElephants	57	0.1	0.8667	-0.0041
1897	HaggardRider	Jess	60	0.084	0.9307	-0.0028
1885	Hammond,William	Astrong-mindedwoman	82	0.08	0.8408	-0.0024
1874	Hardy,Thomas	FarFromtheMaddingCrowd	76	0.088	0.9281	-0.0025
1891	Hardy,Thomas	TessoftheD'Urbervilles	83	0.093	0.9056	-0.0026
1891	Hardy,Thomas	TessoftheD'Urbervilles	83	0.092	0.9177	-0.0022
1895	Hardy,Thomas	JudetheObscure	80	0.086	0.8954	-0.002
1893	Harraden,Beatrice	ShipsThatPassintheNight	20	0.104	0.8539	-0.008
1922	Hart,FrancisRussel	AdmiralsoftheCaribbean	28	0.146	0.8516	-0.0071
1850	Hawthorne,Nathaniel	TheScarletLetter	48	0.114	0.93	-0.004

PubYear	AuthorName	Title	intervals	lexicaldiversity	r2	slope
1851	Hawthorne,Nathaniel	TheHouseoftheSevenGables	59	0.108	0.8654	-0.0028
2007	Hay,Elizabeth	LateNights	55	0.113	0.8808	-0.0034
1796	Hays,Mary	EmmaCourtney	37	0.13	0.9309	-0.0053
1965	Heller,Joseph	Catch22	117	0.089	0.4496	-0.0013
1926	Hemingway,Ernest	TheSunAlsoRises	36	0.08	0.7745	-0.0054
1949	Hemingway,Ernest	ForWhomTheBellTolls	98	0.064	0.6143	-0.0019
1949	Hemingway,Ernest	ForWhomTheBellTolls	94	0.059	0.6904	-0.002
1904	Henry-Ruffin,MargaretEllen	TheNorthStar	45	0.094	0.8521	-0.0029
1919	Hergesheimer,Joseph	JavaHead	39	0.127	0.9526	-0.0044
2008	Hilderbrand,Elin	Barefoot	81	0.092	0.717	-0.0024
1822	Hogg,James	ThreePerilsofMan	114	0.073	0.7561	-0.0017
1885	Hollis,Margery	AnthonyFairfax	22	0.127	0.9318	-0.0047
1894	Hope,Anthony	ThePrisonerofZenda	28	0.104	0.8883	-0.0053
2007	Hosseini,Khaled	AThousandSplendid	59	0.106	0.8472	-0.0027
1885	Howells,WilliamDean	SilasLapham	69	0.069	0.8539	-0.003
1917	Howells,WilliamDean	ConfessionsStAugustine	6	0.251	0.7568	-0.0071
1922	Hudson,WH	ACrystalAge	34	0.111	0.8152	-0.0037
1989	Irving,John	APrayerforOwenMeany	141	0.093	0.7797	-0.0016
2010	Jacobson,Howard	FinklerQuestion	64	0.102	0.7956	-0.003
2011	James,EL	FiftyShadesofGrey	85	0.08	0.7231	-0.003
1881	James,Henry	PortraitofaLady	124	0.054	0.9044	-0.0023
1898	James,Henry	TurnoftheScrew copy	22	0.118	0.9356	-0.005
1903	James,Henry	TheAmbassadors	86	0.07	0.8998	-0.0027
1903	James,Henry	TheAmbassadors copy	86	0.069	0.903	-0.0024
1913	James,Stephens	TheCrockofGold	30	0.117	0.5269	-0.0029
2005	Jin,Ha	WarTrash	74	0.089	0.8432	-0.003
1978	Johnson,Joyce	BadConnections	35	0.135	0.8422	-0.0046
1984	Johnstone,WilliamW	TheLastMountainMan	46	0.085	0.9012	-0.005
1991	Jones,MatthewF	TheCooterFarm	50	0.109	0.8929	-0.0039
1914	Joyce,James	Dubliners	36	0.112	0.8226	-0.0047
1916	Joyce,James	APortraitoftheArtistasaYoungMan	46	0.112	0.3068	-0.0014
1916	Joyce,James	PortraitofanArtist	46	0.112	0.3089	-0.0014
1922	Joyce,James	Ulysses	151	0.115	0.5951	-8e-04
1922	Joyce,James	Ulysses	151	0.115	0.594	-8e-04
1912	Keays,Hersilia	TheroadtoDamascus	52	0.09	0.8967	-0.0045
1891	Kelley,Emma	Megda	50	0.084	0.7992	-0.0035
1957	Kerouac,Jack	OntheRoad	63	0.099	0.8122	-0.0024
2011	King,Stephen	DoctorSleep	95	0.087	0.8531	-0.0024
2014	King,Thomas	TheBackoftheTurtle	63	0.106	0.8356	-0.0033
1849	Kingsley,Charles	AltonLocke	93	0.097	0.7559	-0.0015
1853	Kingsley,Charles	Hypatia	91	0.086	0.9016	-0.002
1988	Kingsolver,Barbara	The Bean Trees	48	0.107	0.7685	-0.0034
1894	Kipling,Rudyard	TheJungleBook	27	0.1	0.7808	-0.0045
1900	Kipling,Rudyard	Kim	58	0.112	0.7295	-0.002
1995	Korman,Keith	SecretDreams	89	0.089	0.8211	-0.0021
2006	Lam,Vincent	BloodlettingandMiraculousCures	47	0.129	0.6064	-0.0021
2006	Larsson,Stieg	TheGirlWhoPlayedWithFire	106	0.073	0.8112	-0.0026
1964	Laurence,Margaret	TheStoneAngel	51	0.119	0.7517	-0.0034
1913	Laut,Agnes	The new dawn	72	0.107	0.8381	-0.0015
1913	Lawrence,D.H.	SonsandLovers	87	0.067	0.938	-0.0031
1928	Lawrence,DH	LadyChatterleysLover	64	0.081	0.8141	-0.0027
1795	Lewis,Matthew	TheMonk	78	0.069	0.8204	-0.0032
1805	Lewis,Matthew	TheBravoofVenice	19	0.144	0.9502	-0.0093
1918	Lewis,Sinclai	TheJob	53	0.144	0.8826	-0.0028
1917	Lewis,Sinclair	TheInnocents	24	0.181	0.7957	-0.0039
1903	London,Jack	TheCalloftheWild	17	0.153	0.8357	-0.0065
1903	London,Jack	CallOfTheWild	17	0.153	0.8556	-0.005
1904	London,Jack	SeaWitch	58	0.101	0.8983	-0.0027
1906	London,Jack	WhiteFang	40	0.102	0.8613	-0.005
1909	London,Jack	MartinEden	77	0.086	0.8718	-0.0022
1987	Longyear,BarryB	SeaOfGlass	56	0.086	0.6866	-0.002
1884	Lyall,Edna	WeTwo	94	0.065	0.8668	-0.0023
2009	Macintyre,Linden	BishopsMan	59	0.101	0.8583	-0.0036
1771	Mackenzie,Henry	TheManofFeeling	20	0.153	0.8513	-0.0062
1855	Manning,Anne	TheoldChelseabun-house	29	0.133	0.8851	-0.0027
1930	Mansfield,Katherine	TheAloe	14	0.172	0.8654	-0.0068
2009	Mantel,Hilary	WolfHall	120	0.076	0.8393	-0.0017
1921	Maquis,Don	TheOldSoak	12	0.17	0.0443	-0.002
1892	Marsham,Clara	CousinSimon	24	0.114	0.9046	-0.0056
1986	Martin,Lori	TheDarklingHills	55	0.087	0.8897	-0.0032
1838	Martineau,Harriet	Deerbrook	122	0.059	0.8341	-0.0019
1919	Maugham,WSomerset	TheMoonandSixpence	41	0.107	0.7586	-0.0037
1865	Mayhew,Augustus	Facesforfortunes	33	0.17	0.8883	-0.002
1851	Melville,Hermann	MobyDick	121	0.094	0.7734	-0.0014
2008	Meyer,Stephanie	TheHost	117	0.065	0.8267	-0.0021
2014	Michaels,Sean	UsConductors	54	0.122	0.7951	-0.003
1989	Miller,GWayne	ThunderRise	79	0.105	0.8445	-0.0025
1992	Mitchell,VE	Imbalance	47	0.113	0.8847	-0.0041
1809	More,Hannah	CoelebsinSearchofaWife	78	0.076	0.8515	-0.0027

PubYear	AuthorName	Title	intervals	lexicaldiversity	r2	slope
2013	Moriarty,Liane	TheHusbandsSecret	63	0.092	0.8817	-0.0038
1985	Morris,Janet	BeyonTheVeil	60	0.115	0.9233	-0.0025
1987	Morrison,Toni	Beloved	54	0.09	0.828	-0.0033
1904	Murfree,MaryNoailles	TheFrontiersman	38	0.15	0.8395	-0.0042
1998	Myers,Tamar	BetweenaWokandaHardPlace	37	0.143	0.917	-0.0041
1958	Nabokov,Vladimir	Lolita	68	0.146	0.8938	-0.0019
1877	Neville,Constance	BehindTheArras	64	0.095	0.9134	-0.0026
1901	Nordau,Max	Thecomedyofsentiment	27	0.131	0.9544	-0.0043
1899	Norris,Frank	McTeague	63	0.086	0.7337	-0.0026
1901	Norris,Frank	TheOctopus	111	0.074	0.7418	-0.0017
1901	Norris,Frank	TheOctopus	111	0.074	0.7762	-0.0015
1903	Norris,Frank	ThePit	69	0.089	0.8796	-0.0029
1952	OConnor,Flannery	WiseBlood	27	0.095	0.7328	-0.006
2009	ONeill,Joseph	Netherland	48	0.155	0.8935	-0.0029
1804	Opie,Amelia	AdelineMowbray	61	0.075	0.9052	-0.0038
1905	Orczy,Emma	TheScarletPimpernel	48	0.094	0.8617	-0.0036
1949	Orwell,George	NineteenEightyFour	59	0.092	0.4507	-0.0028
2012	Otsuka,Julie	BuddhaIn	19	0.157	0.3525	-0.0039
2009	Patterson,James	8thConfession	39	0.14	0.8603	-0.0039
1815	Peacock,ThomasLove	HeadlongHall	16	0.211	0.7514	-0.0077
1818	Peacock,ThomasLove	NightmareAbbey	15	0.185	0.8041	-0.0075
1821	Peacock,ThomasLove	MaidMarian	20	0.156	0.547	-0.0041
1991	Philbrick,Rodman	TheSeventhSleeper	58	0.123	0.8739	-0.0026
2010	Picoult,Jodi	HouseRules	98	0.081	0.9029	-0.0028
1986	Plain,Belva	TheGoldenCup	85	0.084	0.8622	-0.0023
1963	Plath,Sylvia	TheBellJar	39	0.117	0.8863	-0.0056
1838	Poe,EdgarAllen	TheNarrativeofArthurGordonPym	56	0.101	0.2393	-0.0017
2006	Powers,Richard	EchoMaker	97	0.106	0.834	-0.0019
1892	Praed,Campbell	RomannceofachaâÂ¿let	39	0.11	0.7847	-0.004
1993	Proulx,EAnnie	TheShippingNews	62	0.119	0.8146	-0.0024
2009	Pullinger,Kate	MistressofNothing	43	0.093	0.8986	-0.0053
1973	Pynchon,Thomas	GravitysRainbow	198	0.102	0.8421	-9e-04
1790	Radcliffe,Ann	ASicilianRomance	39	0.091	0.8761	-0.0064
1794	Radcliffe,Ann	TheMysteriesofUdolpho	167	0.04	0.8155	-0.0019
1938	Rawlings,MarjorieKinman	TheYearling	69	0.069	0.8754	-0.0033
1903	Read,Opie	TheHarkriders	36	0.112	0.9014	-0.0038
1921	Rhodes,EugeneManlove	StepsonsOfLight	30	0.145	0.8796	-0.0038
2008	Ricci,Nino	OriginofSpecies	101	0.092	0.7616	-0.0019
1915	Richardson,Dorothy	PointedRoofs	32	0.132	0.8702	-0.0052
1902	Rives,Hallie	HeartsCourageous	46	0.139	0.807	-0.0022
1857	Rives,Judith	HomeAndTheWorld	61	0.108	0.8886	-0.0033
2010	Roberts,Nora	SavortheMoment	53	0.106	0.8673	-0.0035
2014	Roberts,Nora	TheCollector	81	0.091	0.754	-0.0023
1886	Roe,Edward	Hefellinlovewithhiswife	56	0.083	0.9008	-0.0028
2007	Roth,Philip	Everyman	21	0.146	0.742	-0.0045
1794	Rowson,Susanna	CharlotteTemple	20	0.124	0.8619	-0.0082
1989	Rule,Jane	AftertheFire	33	0.111	0.9432	-0.0061
1980	Rushdie,Salman	MidnightsChildren	126	0.105	0.8453	-0.0015
1915	Sabatini,Rafael	TheSeaHawk	66	0.092	0.9108	-0.0029
1892	Savage,Richard	ThelittleladyofLagunitas	71	0.111	0.692	-0.0015
1883	Schreiner,Olive	StoryofanAfricanFarm	53	0.076	0.6992	-0.0023
1814	Scott,Walter	Waverley	114	0.085	0.809	-0.0019
1814	Scott,Walter	Waverly	110	0.086	0.8393	-0.0018
1817	Scott,Walter	RobRoy	91	0.098	0.7897	-0.002
1820	Scott,Walter	Ivanhoe	101	0.083	0.7505	-0.0018
1921	Seltzer,Charles	DragHarlan	20	0.128	0.8703	-0.0043
1877	Sewell,Anna	BlackBeauty	30	0.071	0.897	-0.0051
2008	Shaffer,MaryAnn	TheGuernseyLiteraryandPotatoPeelPie	40	0.125	0.6894	-0.0026
1818	Shelley,Mary	Frankenstein	42	0.097	0.8633	-0.0045
1819	Shelley,Mary	Mathilda	20	0.131	0.839	-0.0046
1994	Shields,Carol	TheStoneDiaries	58	0.134	0.8696	-0.0023
1906	Sinclair,Upton	TheJungle	80	0.078	0.6554	-0.0023
1906	Sinclair,Upton	TheJungle	80	0.078	0.6632	-0.002
1917	Sinclair,Upton	KingCoal	67	0.078	0.7987	-0.0026
1983	Singer,Shelley	SamsonsDeal	40	0.102	0.8723	-0.005
2010	Skibsrud,Johanna	Sentimentalists	29	0.108	0.7117	-0.0081
1991	Smiley,Jane	AThousandAcres	72	0.09	0.8107	-0.0027
2000	Smith,Zadie	WhiteTeeth	102	0.103	0.7345	-0.0017
1771	Smollett,Tobias	TheExpeditionofHenryClinker	85	0.101	0.9107	-0.003
1992	Sontag,Susan	TheVolcanoLoverARomance	79	0.103	0.8122	-0.0024
2012	Spalding,Linda	Purchase	53	0.09	0.85	-0.003
2010	Sparks,Nicholas	SafeHaven	64	0.08	0.5368	-0.0028
1906	Stein,Gertrude	ThreeLives	44	0.039	0.7235	-0.0076
1909	Stein,Gertrude	ThreeLives	44	0.039	0.7238	-0.0064
1933	Stein,Gertrude	AutobiographyofAliceBToklas	50	0.073	0.8553	-0.0039
1913	Stevenson,George	Topham'sfolly	47	0.113	0.8483	-0.0028
1882	Stevenson,RobertLouis	TreasureIsland	36	0.096	0.8473	-0.0038
1886	Stevenson,RobertLouis	JekyllandHyde	13	0.162	0.5968	-0.0053
1886	Stevenson,RobertLouis	JekyllandHyde	13	0.162	0.5621	-0.0042

PubYear	AuthorName	Title	intervals	lexicaldiversity	r2	slope
2009	Stockett,Kathryn	TheHelp	88	0.071	0.8418	-0.0026
1897	Stoker,Bram	Dracula	83	0.061	0.8644	-0.0025
1897	Stoker,Bram	Dracula	84	0.065	0.848	-0.002
1997	Stone,Michael	ALongReach	43	0.104	0.8614	-0.0052
1853	Stowe,HarrietBeecher	UncleTom'sCabin	100	0.077	0.8111	-0.0018
1990	Stoyenoff,NormaDavis	OneNighttoRemember	19	0.146	0.9275	-0.0091
1924	Stribling,TS	RedSand	36	0.125	0.8056	-0.0038
1880	Tabor,Eliza	ASylvanQueen	91	0.095	0.8215	-0.0021
1847	Thackeray,William	VanityFair	170	0.063	0.8688	-0.0014
1848	Thackeray,William	VanityFair	170	0.063	0.8651	-0.0014
1857	Trollope,Anthony	BarchesterTowers	108	0.059	0.797	-0.0023
1857	Trollope,Anthony	BarchesterTowers	109	0.06	0.7942	-0.0023
1869	Trollope,Anthony	PhineasFinn	140	0.046	0.8299	-0.002
1875	Trollope,Anthony	TheWayWeLiveNow	190	0.041	0.8332	-0.0017
1876	Trollope,FrancesEleanor	ACharmingFellow	103	0.07	0.8834	-0.0023
1888	Trollope,FrancesEleanor	ThatUnfortunateMarriage	95	0.072	0.8846	-0.0024
1837	Trollope,FrancesMilton	TheVicarofWrexham	95	0.065	0.8359	-0.0024
1855	Trollope,FrancesMilton	TheWidowBarnaby	110	0.055	0.9066	-0.0025
1876	Twain,Mark	TheAdventuresofTomSawyer	38	0.12	0.8929	-0.0037
1876	Twain,Mark	TomSawyer	38	0.116	0.8898	-0.0031
1884	Twain,Mark	TheAdventuresofHuckleberryFinn	56	0.067	0.8902	-0.0036
1884	Twain,Mark	HuckFinn	56	0.066	0.8843	-0.003
1896	Upward,Allen	Acrownofstraw	45	0.106	0.8764	-0.004
1889	Van Dyke,Theodore	RifleRodAndGuninCalifornia	48	0.108	0.4028	-0.0027
2009	Verghese,Abraham	CuttingforStone	118	0.089	0.8489	-0.0016
1969	Vonnegut,Kurt	Slaughterhouse	28	0.143	0.8838	-0.0061
1982	Walker,Alice	TheColorPurple	39	0.082	0.2819	-0.0021
1888	Ward,Mrs.Humphry	RobertElsmere	163	0.066	0.8524	-0.0014
2010	Warren,Dianne	CoolWater	56	0.089	0.9373	-0.0043
1917	Webb,Mary	GonetoEart	45	0.11	0.9404	-0.0038
1893	Wedmore,Frederick	Renunciations	7	0.25	0.7301	-0.0035
1895	Wells,H.G.	TheTimeMachine	17	0.149	0.7821	-0.0058
1918	West,Nathaniel	ReturnOfTheSoldier	16	0.156	0.9013	-0.0047
1905	Wharton,Edith	TheHouseofMirth	73	0.092	0.9023	-0.003
1905	Wharton,Edith	HouseOfMirth copy	73	0.092	0.9014	-0.003
1911	Wharton,Edith	EthanFrome	18	0.137	0.9614	-0.0082
1911	Wharton,Edith	EthanFrome	18	0.136	0.9575	-0.0071
1913	Wharton,Edith	CustomoftheCountry	78	0.089	0.9073	-0.0023
1920	Wharton,Edith	TheAgeofInnocence	57	0.102	0.8896	-0.0033
1920	Wharton,Edith	AgeofInnocence	58	0.101	0.8732	-0.0028
1890	Wilde,Oscar	ThePictureofDorianGray	42	0.092	0.7748	-0.0041
1890	Wilde,Oscar	PictureofDorianGrey	42	0.093	0.7768	-0.0042
1914	Williams,Irving	BigWallace	22	0.141	0.7726	-0.0044
1994	Willis,Connie	UnchartedTerritory	22	0.117	0.8665	-0.0085
1982	Wilson,Steve	DealersWheels	33	0.132	0.8918	-0.0052
1894	Wolf,Emma	Aprodigalinlove	39	0.122	0.8173	-0.0037
1987	Wolfe,Tom	TheBonfireoftheVanities	147	0.086	0.7999	-0.0016
1788	Wollstonecraft,Mary	Mary	13	0.176	0.9016	-0.0085
1798	Wollstonecraft,Mary	Maria	25	0.141	0.9204	-0.0057
1922	Woolf,Virginia	JacobsRoom	31	0.146	0.8254	-0.0028
1925	Woolf,Virginia	Mrs.Dalloway	35	0.12	0.8085	-0.0044
1925	Woolf,Virginia	MrsDalloway	35	0.12	0.8083	-0.0044
1927	Woolf,Virginia	TotheLighthouse	38	0.104	0.7711	-0.0048
1927	Woolf,Virginia	ToTheLighthouse	38	0.104	0.771	-0.0048
1928	Woolf,Virginia	Orlando	43	0.123	0.8841	-0.0036
2008	Wroblewski,David	TheStoryofEdgarSawtelle	107	0.073	0.7712	-0.0023
1844	Yonge,Charlotte	Abbeychurch	40	0.088	0.7459	-0.0049
1850	Yonge,Charlotte	Henrietta'sWish	48	0.082	0.9209	-0.0041
1853	Yonge,Charlotte	TheHeirofRedcliffe	128	0.052	0.8956	-0.0021
1979	Yorgason,BlaineM	TheWindwalker	11	0.116	0.6665	-0.0076
2007	Young,WilliamP	TheShack	44	0.105	0.9351	-0.0048